Sygnatura: Pol J Radiol, 2005; 70(2): 25-30

Polish Journal of Radiology
www.PolRadiol.com
ORIGINAL ARTICLE

Otrzymano: 2004.11.15
Zaakceptowano: 2004.12.10

Evaluation of alternative methods for coronary calcium scoring in multi-detector-row computed tomograhy

Ocena alternatywnych metod pomiaru ilości zwapnień w tętnicach wieńcowych w wielorzędowej tomografii komputerowej

Zbigniew Serafin, Władysław Lasek, Katarzyna Laskowska, Małgorzata Marzec

Department of Radiology and Diagnostic Imaging The Ludwik Rydygier Medical University in Bydgoszcz, Poland

Adres autora: Department of Radiology and Diagnostic Imaging The Ludwik Rydygier Medical University in Bydgoszcz, ul. M. Curie-Skłodowskiej 9, 85-094 Bydgoszcz, e-mail: biuro@radiologia.bydgoszcz.pl

Summary

Background: Determining coronary artery calcifications is one of the methods of coronary heart disease screening. However, the traditional Agatston Calcium Score (CS) shows low interexamination reproducibility. The aim of this study was to evaluate the interscan variability coefficients of calcium measures based on three modifications of the original Agatston equation.

Material/Methods: Fifty adults (37 men and 13 women; mean age 46.2 ± 9.2 years) were included in the study. Each patient was examined with two consecutive, prospectively electrocardiographically triggered, multi-detector-row CT acquisitions to detect and quantify coronary artery calcifications. CS was calculated according to the method by Agatston et al. Alternative scores were calculated using a continuous weighting factor (CS-CM), the average lesion attenuation value (CS-SA), or both (CS-CA). The mean and median interscan percent variabilities of the methods were evaluated using nonparametric analysis of variance.

Results: In the 50 patients, 1315 calcified lesions were found. The alternative scores correlated well with CS (for CS vs. CS-SA, CS-CM, and CS-CA, $r = 0.990$, 0.840, and 0.946, respectively, $P < 0.0001$). The mean and median percent variabilities did not differ significantly among the methods tested ($P = 0.370$). For CS, CS-SA, CS-CM, and CS-CA the mean variabilities were 13.24%, 13.36%, 16.00%, and 13.62%, respectively. Except for CS-CM, the methods showed similar distributions of variability vs. score on Bland and Altman plots.

Conclusion: None of the tested modifications of the Agatston method brought improvement in the interscan reproducibility of coronary calcium scoring. In our opinion, a significant reduction in variability may be achieved by a standardization of image acquisition and reconstruction.

Key word: computed tomography • atherosclerosis • coronary artery calcification

PDF file: http://www.polradiol.com/pub/pjr/vol_70/nr_2/6599.pdf

Background

Coronary calcium measurement could be an efficient, non-invasive test for coronary heart disease (CHD) screening and monitoring. Several papers report that calcium score is a strong and independent risk factor for CHD [1–3]. Grundy proposed that coronary calcium measurement might replace age as a risk factor in global CHD risk assessment [4]. However, the classic Calcium Score (CS) [5] was reported to have limited interexamination reproducibility, with a measured variability of between 14 and 51% [6–8]. Clinical purposes require that the method have an error significantly lower than the expected change in calcified plaque burden over time. According to numerous authors, the average

Table 1. Summary statistics of the four methods tested, based on the results from the first acquisition (50 patients, 1315 lesions).
Tabela 1. Statystyki opisowe badanych metod w oparciu o wyniki pierwszej akwizycji (50 pacjentów, 1315 zmian uwapnionych).

	CS	CS-SA	CS-CM	CS-CA
Mean score	704.12	424.91	999.22	456.14
Median score	341.35	194.85	326.65	208.55
Absolute SD*	794.097	477.052	1270.772	512.744
Relative SD*	1.1278	1.1227	1.2718	1.1241
Minimal score	5.3	4.1	5.6	4.5
Maximal score	2952.3	1831.9	5836.7	1818.0

* Standard Deviation

annual increase in CS reaches 14–155% [9–12]. On the other hand, effective statin therapy may reduce calcifications by 7% a year on average [13]. Thus the variability of the calcium measurement should not exceed 10% for the reliable detection of small lesions with a reasonable follow-up interval [14].

Since the publication of Agatston et al. [5] there have been many modifications of the method of coronary calcium calculation. These include different CT attenuation thresholds [9, 15], measurement of area [16], volume [11, 14, 17], and mass [14, 15, 18] of the calcifications, the use of different scan protocols [14, 15, 18], as well as some modifications of the classic CS equation. The main aim of these modification was to improve the interexamination variability of measurements.

The purpose of this paper was to evaluate the interscan variability and correlations among alternative methods of coronary Calcium Score measurement.

Materials and Methods

Patient population

Fifty patients were included with Agatston Calcium Scores over 0 and with suspected or established CHD. There were 37 men and 13 women aged 30–63 years (mean age 46.2 ± 9.2 years). The study was approved by the Board of Bioethics of our university. Exclusion criteria were: age less than 30 years, bypass surgery, coronary stent placement, dyspnoe, and heart rate over 90 beats per minute (bpm). The mean heart rate during scanning was 71 ± 14 bpm.

Image acquisition

A four-row CT scanner (Mx8000, Philips Medical Systems, Cleveland, Ohio) was used. Each patient was scanned twice with a five-minute interval. The patients remained stationary on the scanner table between the acquisitions. The examination was planned separately for the consecutive acquisitions with a small movement of the starting point (1–3 mm). Scans were obtained during one held breath from a level approximately 0.5 cm below the carina to the diaphragm. The scan parameters were: average number of slices: 48, scan time: 0.33 s, gantry rotation time: 0.5 s, collimation: 4 × 2.5 mm, filter B (normal), matrix: 512, 120 kVp, 165 mAs, and axial prospective electrocardiographic gating at 60%.

Calcium measurements and statistical analysis

The examinations were evaluated by one investigator using version 5.0 of *Heart Beat-CS* software on an *MxView* workstation (Philips Medical Systems, Cleveland, Ohio). Structures with CT attenuation above the threshold of 130 HU were automatically distinguished. Coronary calcifications were then manually localized and calcium scores were calculated according to the chosen parameters. The traditional Calcium Score (*CS*) was calculated according to the method by Agatston et al. [6] as the sum of the area a multiplied by the weighting factor F of n lesions:

$$CS = \sum_{n} CS(n) = \sum_{n} a(n) \cdot F(n)$$

The factor F depends on the maximum CT number (CT#) of the lesion: $F = 1$ if $130 \leq$ CT# < 200 HU, $F = 2$ if $200 \leq$ CT# < 300 HU, $F = 3$ if $300 \leq$ CT# < 400 HU, and $F = 4$ if CT# ≥ 400 HU.

Additionally, three alternative calcium scores were evaluated. In two of them the traditional stepwise weighting factor was replaced by a continuous weighting factor Fc, which was calculated by the formula:

$$F_c(n) = \frac{{}_{max}CT\#(n)\text{-}50}{100}$$

where ${}_{max}CT\#$ is the peak CT value of the lesion and n the number of lesions. For CT values below 450, this method gives similar results to the stepwise method. In the center of each range, i.e. for CT# 150, 250, 350, and 450 HU, the results are the same as in the stepwise method. For CT# above 450, where the weight according to the step method is 4, the continuous method results continue to increase. In two methods, both weighting factors, F and F_c, were calculated using the average CT# instead of the maximum CT# of the above-threshold area. The alternative scores were called: (a) CS-SA (stepwise F and average CT#); (b) CS-CM (continuous F and maximum CT#); and (c) CS-CA (continuous F and

Table 2. Statistical validation of the four algorithms based on interexamination variability.
Tabela 2. Statystyczna walidacja badanych algorytmów na podstawie zmienności pomiarów powtarzanych.

	CS	CS-SA	CS-CM	CS-CA
Mean relative variability	13.24%	13.36%	16.00%	13.62%
Median relative variability	7.82%	7.75%	8.50%	7.87%
Absolute SD*	14.24%	16.06%	15.19%	16.15%
Minimal variability	0.11%	0.00%	0.10%	0.30%
Maximal variability	66.67%	69.36%	56.44%	78.19%
*d***	-1.3%	1.1%	1.6%	2.1%

* standard deviation
** systematic error, mean difference as a percentage of the average of the two consecutive measurements

average CT#). In this manner the conventional Calcium Score could be called CS-SM (stepwise *F* and maximum CT#).

The basic marker used for determining the accuracy of the methods was the relative difference between two acquisitions (percent variability):

$$v = \frac{2 \times |x_1 - y_1|}{(x_1 + y_1)} \times 100\%$$

where x_i and y_i were the scores from the two consecutive scans of the same lesion i. Statistical association among the calcium scores was evaluated with linear regression equations and Pearson correlation coefficients. The relationship between variability and calcium distribution was shown by Bland-Altman plots; for this purpose, the raw data were log-transformed ($\log_{10}$[mean score]) to reduce skewness [19]. The systematic error (d) and the limit of agreement ($d \pm 1.96\ SD$, where SD is the standard deviation) of the consecutive measurements were determined according to the method of Bland and Altman [20]. Results from the two acquisitions within the scoring method were compared using the Wilcoxon matched-pairs test. To test differences among the mean variabilities of the four scores, the Friedman test was used. In the statistical analysis a P value less than 0.05 was established as indicating a statistically significant difference.

Results

In the examined group of 50 patients, 1315 calcified lesions were found. Based on the first scan, mean and median CS were 704.12 and 341.35, respectively, with a standard deviation of 794.1. Two patients were described as having low-grade calcifications (CS 1–10), 12 mild calcifications (CS 11–100), 12 moderate (CS 101–400), and 24 severe calcifications (CS more than 400). The lowest measured CS was 5.3 and the highest 2952.3. The traditional score from the two acquisitions had very good correlation ($r = 0.995, P < 0.0001$). Results of calcium scoring using each method are presented

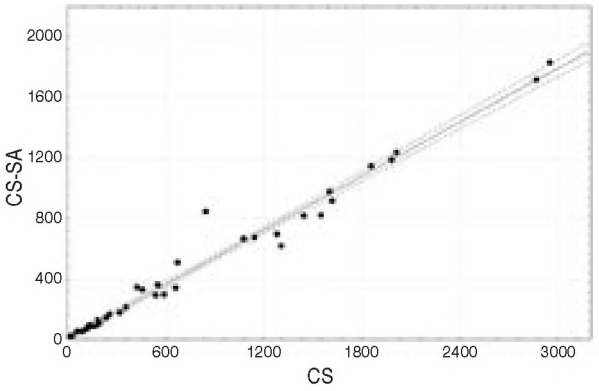

Figure 1. Scatterplot shows relationship between CS results and CS-SA results, based on the first scan. Regression equation: CS-SA = 5.88 + 0.60 · CS; $r = 0.990$, $P < 0.0001$. Black line: regression line, dashed line: 95% confidence interval.
Rycina 1. Wykres zależności między wynikami CS i CS-SA uzyskanymi w pierwszej akwizycji. Wykres regresji: CS-SA = 5,88 + 0,60 · CS; $r = 0.990$, $P < 0.0001$. Linia czarna – linia regresji, linia przerywana – 95% przedział ufności.

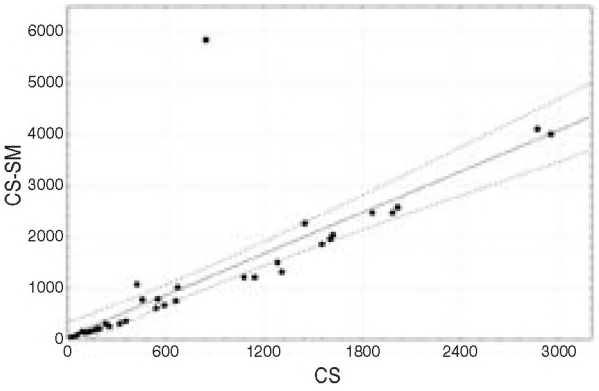

Figure 2. Scatterplot shows relationship between CS results and CS-CM results, based on the first scan. Regression equation: CS-CM = 52.45 + 1.34 · CS; $r = 0.840$, $P < 0.0001$. Black line: regression line, dashed line: 95% confidence interval.
Rycina 2. Wykres zależności między wynikami CS i CS-CM uzyskanymi w pierwszej akwizycji. Wykres regresji: CS-CM = 52,45 + 1,34 · CS; $r = 0.840$, $P < 0.0001$. Linia czarna – linia regresji, linia przerywana – 95% przedział ufności.

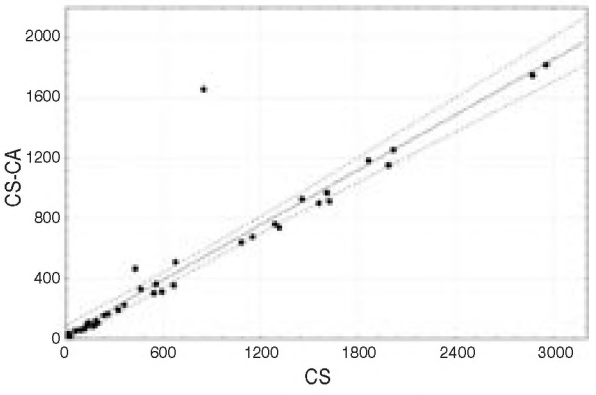

Figure 3. Scatterplot shows relationship between CS results and CS-CA results, based on the first scan. Regression equation: CS-CA = 26.01 + 0.61 · CS; $r = 0.946$, $P < 0.0001$. Black line: regression line, dashed line: 95% confidence interval.

Rycina 3. Wykres zależności między wynikami CS i CS-CA uzyskanymi w pierwszej akwizycji. Wykres regresji: CS-CA = 26,01 + 0,61 · CS; $r = 0.946$, $P<0.0001$. Linia czarna – linia regresji, linia przerywana – 95% przedział ufności.

in Table 1. The alternative scoring methods correlated very well with CS. Correlation coefficients for CS vs. CS-SA, CS-CM, and CS-CA were 0.990, 0.840, and 0.946, respectively ($P < 0.0001$) (Figures 1–3).

Results from the two acquisitions did not differ statistically for all the methods ($P = 0.2775$–0.6629, Wilcoxon matched-pair test). The mean percent variabilities of the methods were 13.24% for CS, 13.36% for CS-SA, 16.00% for CS-CM, and 13.62% for CS-CA (Table 2). The median percent variabilities were 7.82%, 7.75%, 8.50%, and 7.87%, respectively (Figure 4). The differences between the values of mean percent variability of the four scores were not statistically significant (χ^2 *ANOVA* = 3.144, $P = 0.370$, Friedman test). Systematic error of the algorithms reached similar values, from -1.3% to 2.1%. Bland and Altman plots (Figure 5) present the distribution of variability vs. the score.

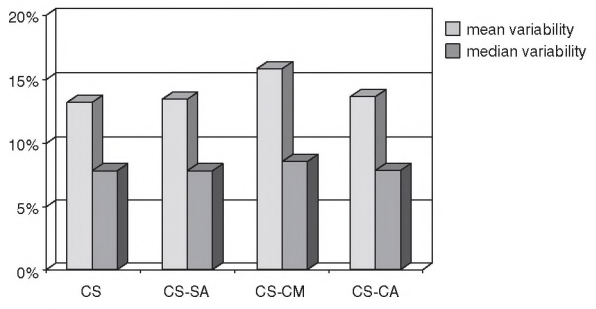

Figure 4. Comparison of mean and median relative interscan variability of the calcium scoring methods. The differences were not statistically significant according to the Friedman test.

Rycina 4. Porównanie średniej i mediany względnej zmienności pomiarów powtarzanych badanych metod obliczania wskaźnika zwapnień. Różnice nie były statystycznie istotne w teście Friedmana.

Discussion

Sufficient reproducibility of coronary calcium measurements is a key condition to establish it as a standard clinical test for coronary heart disease risk determination. In view of studies of Callister et al. [13] and Budoff et al. [21], who showed a significant reduction in the rate of coronary calcium progression or even a decrease in calcium amount in patients on statin therapy, calcium quantification can be used as a test for therapy monitoring as well.

Many authors have criticized the original Agatston method for its low interexamination reproducibility [7, 11, 15]. The substantial variability of the results was suspected to be produced mainly because of the stepwise calculation of the weighting factor F based on the peak attenuation values of a plaque. In this view the Agatston Score was called a semiquantitative method [15]. Although several modifications of both the acquisition method and the measurement algorithms were proposed [9, 11, 14, 15, 17, 18], large cross-sectional studies, including MESA [22] and MUNICH [23], are rather conservative in this matter, using the Agatston score as a measure. The reason may be the uncertain advantages of those modifications [50]. Moreover, most of the studies reporting the clinical value of coronary calcium measurement were performed using the traditional calcium score [1–3, 9, 24, 25].

In our material the conventional calcium score measurement resulted in interscan reproducibility comparable to those of Hong et al., who determined calcium volume and absolute calcium mass [18] (13.2% vs. 13.9% and 9.3%, respectively). The variability of the Agatston score in their study was 20.4%. However, since the reproducibility is strongly dependent on lesion size due to partial volume averaging, it is difficult to compare these results directly. In future studies, the relation between different calcium measures should be performed with respect to calcium dispersion. Thus, such a low variability of the traditional calcium score in our material may be a result of the relatively large lesions seen in our subjects.

It was proposed that the reproducibility of calcium scoring may be improved by adopting a more continuous rather than a step-wise measurement method [26, 27]. In this study we assessed the value of three modifications of the Agatston calcium scoring method approaching this idea. In CS-SA method, the weighting factor was calculated using the average instead of the peak lesion density, CS-CM included a continuous weighting factor, and CS-CA was a compilation of those. The results calculated with these methods correlated very well with the traditional calcium score. Since all the scores presented a significant dispersion of the variability results (0%–78%), differences between the relative interscan variabilities of the calcium score and its modifications were not statistically significant in the Friedman test (when compared method-by-method by the Wilcoxon matched-pair test, only the difference in variability of CS-CM vs. CS-CA was at the limit of significance, $P = 0.05$). Our results indicate that in a larger sample of patients taken from the same population, the differences between CS-CM and the other scores could reach statistical significance. Our results should be validated in large

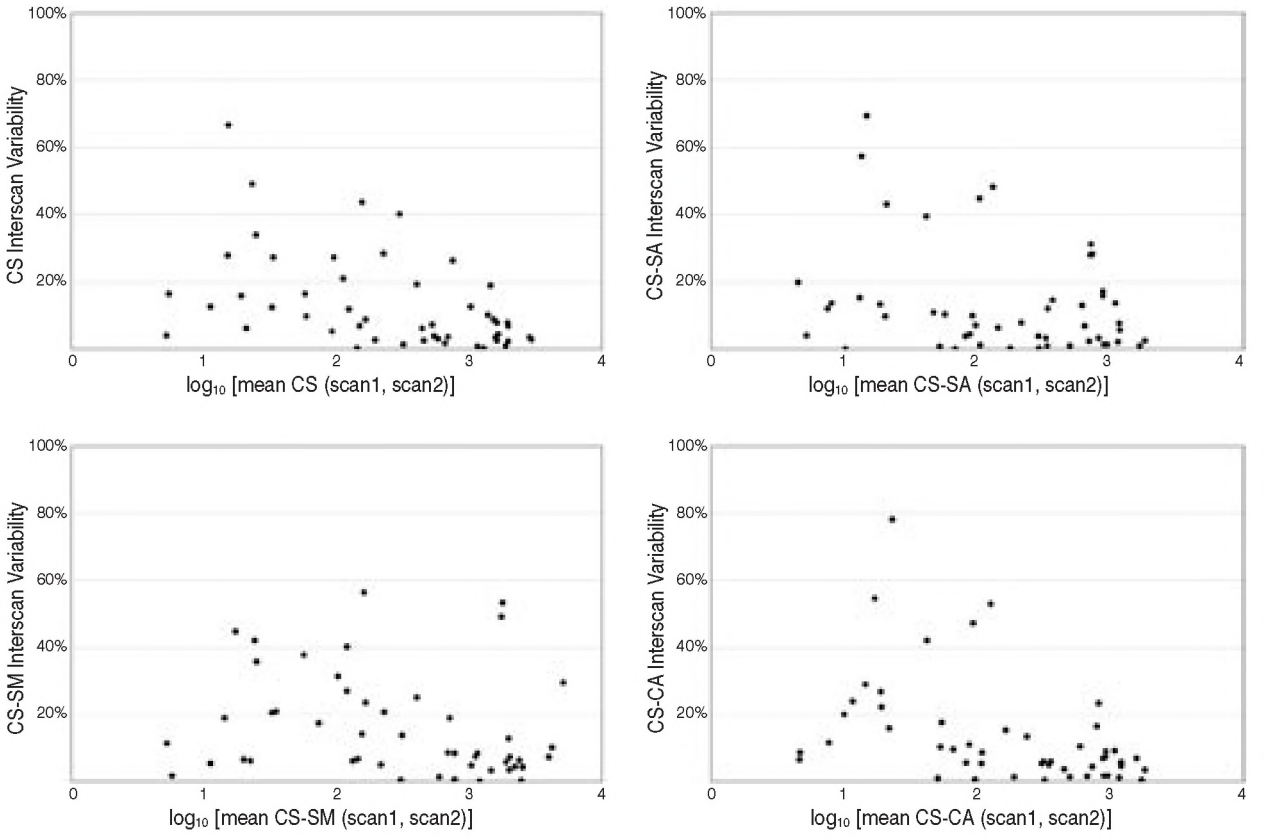

Figure 5. Bland and Altman plots show changes in the percent interexamination variability with increasing mean score. CS-CM demonstrates absolutely systematic error. The variability of CS, CS-SA, and CS-CA depends on the magnitude of the measurement (proportional error). To reduce the skewness of the data, the scores were log10 transformed.

Rycina 5. Wykresy Blanda-Altmana pokazują zmiany procentowej zmienności pomiarów powtarzanych wraz ze wzrostem średniego wskaźnika zwapnień. CS-CM charakteryzuje się błędem całkowicie systematycznym. Zmienność CS, CS-SA i CS-CA zależy od wielkości pomiaru (błąd proporcjonalny). W celu zmniejszenia skośności danych wskaźniki poddano logarytmowaniu dziesiętnemu.

cohorts in a future study. Bland and Altman plots also show a similar type of variability for CS, CS-SA, and CS-CA (proportional error), while the plot for CS-CM indicates an absolutely systematic error.

Overall, the Agatston score showed the best reproducibility of the methods tested, and in view of these results the proposed modifications of the traditional calcium scoring seem to be ineffective in improving score reproducibility. Shemesh et al. reached a 31% decrease (from 32% to 23%) in interscan variability using the average algorithm on double helical CT [26]. In our study, CS-SA had significantly lower variability (13.36%) and did not differ from the conventional score. Some factors in our study might contribute to this lower variability, such as a more advanced CT scanner, a different scanning protocol, and a potentially different pattern of calcifications in our group of patients.

Except for CS-CA, the tested algorithms remain semiquantitative in their nature and have some of the disadvantages of the Agatston method. In our material, the interscan reproducibility of CS-CA was not significantly improved compared with the Agatston score. Thus we could assume that the stepwise calculation of the traditional calcium score is not the source of its limited reproducibility.

Most of the recent papers concerning validation of coronary calcium measurements with multi-detector-row CT were published based on studies using scanners from one manufacturer, Siemens Medical Systems. The unique technical and software solutions introduced by the scanners' producers present an important problem of potential incompatibility of calcium scoring results. While the variability is still not satisfactory according to Ohnesorge at al. [14], there is the question whether one should concentrate on improving reproducibility on a single scanner rather than on turning to a method independent of the device. In our opinion the second way is the more promising, and especially calcium absolute mass quantification is very accurate, since it eliminates the influence of different scanner protocols and properties by using phantom calibration [15].

References:

1. Shaw LJ, Raggi P, Schisterman E, Schisterman E, Berman DS, Callister TQ. Prognostic value of cardiac risk factors and coronary artery calcium screening for all-cause mortality. Radiology 2003; 228: 826–833.
2. Raggi P, Cooil B, Callister TQ. Use of electron beam tomography data to develop models for prediction of hard coronary events. Am Heart J 2001; 141: 375–382.
3. Taylor AJ, Feuerstein I, Wong H, Barko W, Brazaitis M, O'Malley PG. Do conventional risk factors predict subclinical coronary artery disease? Results from the Prospective Army Coronary Calcium Project. Am Heart J 2001; 141: 463–468.
4. Grundy SM. Coronary plaque as a replacement for age as a risk factor in global risk assessment. Am J Cardiol 2001; 88: 8E–11E.
5. Agatston AS, Janowitz WR, Hildner FJ, et al. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 1990; 15: 827–832.
6. O'Rourke RA, Brundage BH, Froelicher VF, et al. American College of Cardiology/American Heart Association Expert Consensus Document on Electron-Beam Computed Tomography for the Diagnosis and Prognosis of Coronary Artery Disease. Circulation 2000; 102: 126–140.
7. Wang S, Detrano RC, Secci A, et al. Detection of coronary calcification with electron-beam computed tomography: evaluation of interexamination reproducibility and comparison of three image-acquisition protocols. Am Heart J 1996; 132: 550–558.
8. Bielak LF, Sheedy PF, Peyser PA. Coronary artery calcification measured at electron-beam CT: agreement in dual scan runs and change over time. Radiology 2001; 218: 224–229.
9. Shemesh J, Apter S, Stroh CI, Itzchak Y, Motro M. Tracking coronary calcification by using dual-section spiral CT: a 3-year follow-up. Radiology 2000; 217: 461–465.
10. Maher JE, Bielak LF, Raz JA, et al. Progression of coronary artery calcification: a pilot study. Mayo Clin Proc 1999; 74: 374–355.
11. Callister TQ, Cooil B, Raya SP, Lippolis N, Russo DJ, Raggi P. Coronary artery disease: improved reproductibility of calcium scoring with an electron-beam CT volumetric method. Radiology 1998; 208: 807–814.
12. Yoon H-C, Emerick AM, Gjertson DW, Goldin JG. Calcium begets calcium: progression of coronary artery calcification in asymptomatic subjects. Radiology 2002; 224: 236–241.
13. Callister TQ, Raggi P, Cooil B, Lippolis NJ, Russo DJ. Effect of HMG-CoA reductase inhibitors on coronary artery disease as assessed by electron-beam computed tomography. N Engl J Med 1998; 339: 1972–1978.
14. Ohnesorge B, Flohr T, Fischbach R, et al. Reproducibility of coronary calcium quantification in repeat examinations with retrospectively ECG-gated multisection spiral CT. Eur Radiol 2002; 12: 1532–1540.
15. Hong C, Becker CR, Schoepf UJ, et al. Coronary artery calcium: absolute quantification in nonenhanced and contrast-enhanced multi-detector row CT studies. Radiology 2002; 223: 474–480.
16. Kaufmann RB, Peyser PA, Sheedy PF, et al. Electron beam computed tomographic coronary calcium score sutpoints and severity of associated angiographic lumen stenosis. J Am Coll Cardiol 1997; 25: 626–632.
17. Ohnesorge B, Becker CR, Flohr TG, et al. Multi-slice CT in cardiac imaging. Technical principles, clinical applications and future developments. Berlin, Springer-Verlag, 2002.
18. Hong C, Bae KT, Pilgram TK. Coronary artery calcium: accuracy and reproducibility of measurements with multi-detector row CT – assessment of effects of different thresholds and quantification methods. Radiology 2003; 227: 795–801.
19. Lu B, Zhuang N, Mao S-S, et al. EKG-triggered CT data acquisition to reduce variability in coronary arterial calcium score. Radiology 2002; 224: 838–844.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–310.
21. Budoff MJ, Lane KL, Bakhsheshi H, et al. Rates of progression of coronary calcium by electron beam tomography. Am J Cardiol 2000; 86: 8–11.
22. Bild DE, Bluemke DA, Burke GL, et al. The Multi-Ethnic Study of Atherosclerosis: Objectives and Design. Am J Epidemiol 2002; 156: 871–881.
23. Schmermund A, Erbel R, Silber S. Age and gender distribution of coronary artery calcium measured by four-slice computed tomography in 2030 persons with no symptoms of coronary artery disease. Am J Cardiol 2002; 90: 168–173.
24. Yoon H-C, Emerick AM, Hill JA, Gjertson DW, Goldin JG. Calcium begets calcium: progression of coronary artery calcification in asymptomatic subjects. Radiology 2002; 224: 236–241.
25. Kennedy J, Shavelle R, Wang S, Budoff M, Detrano RC. Coronary calcium and standard risk factors in symptomatic patients referred for coronary angiography. Am Heart J 1998; 135: 696–702.
26. Shemesh J, Tenenbaum A, Kopecky KK, et al. Coronary calcium measurements by double helical computed tomography. Using the average instead of peak density algorithm improves reproducibility. Invest Radiol 1997; 32: 503–506.
27. Yoon HC, Greaser LE, Mather R, Sinha S, McNitt-Gray MF, Goldin JG. Coronary artery calcium: alternate methods for accurate and reproducible quantification. Acad Radiol 1997; 4: 666–673.